基于用户关系的跨社交网络用户身份关联方法 *

刘奇飞 [a]，杜彦辉 [a, b]，芦天亮 [a, b†]

(中国人民公安大学 a. 信息技术与网络安全学院; b. 网络空间安全与法治协同创新中心, 北京 100038)

摘 要：为识别出不同社交网络平台中属于同一自然人的账号，提出了一种基于用户关系的跨社交网络用户身份关联方法。首先，设计了基于网络表示学习的用户关系提取模块，将大规模用户关系转换至低维向量空间进行表示；然后，针对异构信息网络改进了传统网络表示学习算法，提出了 CSN_LINE 算法，实现融合跨社交网络先验关联关系的网络表示；最后，构建了基于多层感知机的用户身份关联模型。实验结果表示，提出的方法与目前先进的方法相比，综合指标 F1 值和正确率的提高均超过 12%，证明了该方法的合理性和有效性。

关键词：用户关系；跨社交网络；用户身份关联；网络表示学习；多层感知机

中图分类号：TP391 doi: 10.19734/j.issn.1001-3695.2018.07.0517

User identity linkage across social networks based on user relations

Liu Qifei[a], Du Yanhui[a, b], Lu Tianliang[a, b†]

(*a. Information Technology & Network Security Institute, b. CIC of Security & Law for Cyberspace, People's Public Security University of China, Beijing* 100038, *China*)

Abstract: In order to distinguish the accounts that belong to the same person, this paper proposed a method to link user identity across social networks based on user relations. Firstly, we designed a user relations feature extraction module based on network representation learning. It could embed large information networks into low-dimensional vector spaces. Secondly, we proposed CSN_LINE algorithm for heterogeneous information network. The improved algorithm could represent network combining with anchor links across networks. Finally, we constructed a user identity linkage model based on multi-layer perception . Experiments showed that the F1 rate and accuracy rate of this method increased over 12% compared with the current advanced algorithm. The validity and rationality of the method is proved.

Key words: user relations; across social networks; user identity linkage; network representation learning; multi-layer perception

0 引言

Globalwebindex 公司最近的研究表明，就全球范围来看，98%的网络用户至少使用了一个社交网络，平均每个网络用户拥有 7.6 个网络账户。网民普遍拥有多个社交网络平台的身份，用户信息分散在各个不同的社交网络平台上，为了打破这种“信息孤岛”现象，实现多源异构数据融合，进行跨社交网络用户身份关联是十分关键。用户身份关联可以为复杂的社交网络分析业务提供更丰富的数据支撑。例如刻画出更全面的用户画像，帮助商业推荐系统为用户提供更精准的个性化服务，解决推荐系统“冷启动”难题，也能在网络安全领域为识别虚假账号、非法账号提供支持，拥有广泛的应用价值，所以跨社交网络用户身份关联是十分有意义的研究课题。

目前有许多研究关注于通过用户属性信息和用户行为信息进行用户身份关联，也取得了一些成果。但是由于目前隐私保护越来越受到重视，用户属性和用户行为信息难以获取，且难以确认信息的真实性，这给用户身份关联带来了极大的挑战。为了克服这一困难，充分利用用户关系是很有意义的，用户关系拓扑结构具有匿名性，同时一个普通的网民用户不会在个人账户中虚假地添加一些无意义的关联关系，用户关系更真实地展现一个用户的实际情况，体现着一个用户的情感、兴趣，反映一个自然人在现实世界中的社会关系，基于用户的社交网络关系实现跨社交网络用户身份关联可以弥补基于用户属性和用户行为的一些不足，提高跨社交网络用户身份关联方法的鲁棒性和泛化能力。

研究基于用户关系的跨社交网络用户身份关联问题也面临着许多难题：a)用户关系难以进行定量表示，将用户关系进行网络表示并反映拓扑结构特征比较困难;b)社交网络平台的用户是海量的，如何在大规模复杂网络中实现高效的多账号关联算法是一个难点问题;c)由于社交网络的无标度性和小世界性，用户的关系拓扑结构存在高度的同质性，难以通过算法达到精准的用户身份关联效果。

针对问题，本文提出了一种基于用户关系的跨社交网络用户身份关联方法，主要工作如下：a）利用网络表示学习，设计了用户关系的特征提取方法，将用户关系特征转换至低维向量空间中进行表示;b）针对关联同一自然人在不同社交网络平台账号这一应用场景，面向此类异构信息网络，改进了传统网络表示学习算法，提出了 CSN_LINE;c）基于多层感知机，设计了基于用户关系的跨社交网络用户身份关联模型;d）面向新浪微博和豆瓣，获取了大规模用户关系数据，对本文设计的方法进行了训练和验证。

收稿日期：2018-07-08；**修回日期：**2018-08-11 **基金项目：**国家重点研发计划重点专项项目（2017YFB0802804）；国家自然科学基金资助项目（61602489）；中国人民公安大学 2018 年基本科研业务费科研机构项目（2018JKF504）

作者简介：刘奇飞（1994-），男，硕士研究生，主要研究方向为社交网络、机器学习；杜彦辉（1969-），男，教授，博士，主要研究方向为信息安全；芦天亮（1985-），男（通信作者），副教授，博士，主要研究方向为信息安全（lutianliang@ppsuc.edu.cn）.

1 相关工作

研究者通常基于用户属性、用户行为、用户关系三个不同的维度设计跨社交网络用户身份关联方法。其中用户属性特征主要包括用户名、个人描述、性别、职业、头像等[1~4]，用户行为特征主要包括用户发布内容的文体风格[5~7]、用户行动轨迹[8, 9]等。在用户关系方面，也有许多研究者开展了相应的研究。

Liu 等人[10]通过长期行为分析和短期多角度信息匹配来为用户行为建模，同时运用用户的 ego-network 的结构同构性，提出 HYDRA 多账号关联方法。Tan 等人[11]利用超图将网络关系表示为矩阵，同时通过降低矩阵的维度来减少关联算法的计算量。一般而言，用户关系是可以通过邻接矩阵进行表示，但在大规模网络中这个矩阵比较稀疏。Man 等人[12]提出了 PALE 方法，利用网络表示学习将用户节点映射到低维向量空间进行表示，再利用关联模型实现用户身份关联。Feng 等人[13]设计了两种新的衡量不同社交网络用户间相似度的方法。Zhang 等人[14]提出了 COSNET 方法，综合社交网络拓扑图的局部匹配信息和全局匹配信息，利用能量模型来解决多账号关联问题。Zhou 等人[15]提出了 FRUI 方法，充分利用已关联的跨社交网络用户对，大大降低时间复杂度。汪潜等人[16]利用众包的方式增加训练样本的数据量，然后运用全视角的特征来衡量用户之间的相似度，提出了一种基于全视角特征结合众包的跨社交网络用户识别方法。

2 问题分析

基于用户关系的用户身份关联的目的是利用用户关系识别出同一自然人在不同社交网络平台的账号，对本问题进行形式化描述如下。

存在两个不同的社交网络平台，分别为 G_A 和 G_B，G_A=（V^A，E^A），G_B=（V^B，E^B），其中 V^A 和 V^B 表示社交网络平台 A 和 B 中的用户集合，E^A 和 E^B 表示社交网络平台 A 和 B 中的用户关系集合，跨社交网络关联关系为 M，M={（v，u）|$v\in V^A$ ，$u\in V^B$ }，集合 M 包含不同社交网络平台中属于同一自然人的用户对。

如图 1 所示，在社交网络平台 A 和 B 中，平台内部用户之间存在一些关联关系（图 1 中实线），例如关注关系、好友关系等，同时在不同平台之间也存在一些先验关联关系（图 1 中上方 3 条虚线），即预先确定的属于同一自然人的用户对之间的关系。利用上述平台内部关系和平台之间的关系，基于用户关系的用户身份关联需要识别出更多未被发现的跨社交网络关联关系（图 1 中下方两条虚线）。

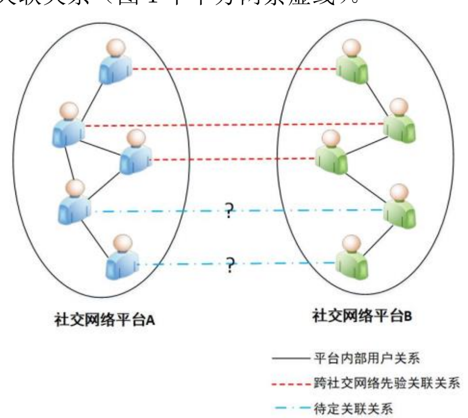


图 1 基于用户关系的用户身份关联问题

Fig. 1 User identity association based on user relations

3 跨社交网络用户身份关联方法

本文提出的基于用户关系的跨社交网络用户身份关联方法主要包括两个关键部分：用户关系特征提取模块和基于多层感知机的用户身份关联模型，本方法的流程图如图 2 所示。

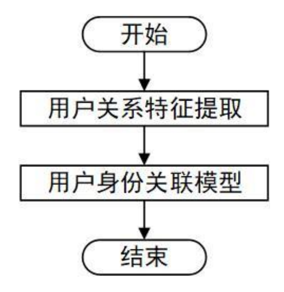


图 2 用户身份关联方法流程图

Fig. 2 Flow of user identity association method

其中用户关系特征提取模型实现了用户节点网络拓扑信息的向量化，将用户节点拓扑特征通过低维稠密的向量进行表示。基于多层感知机的用户身份关联模型是利用多层感知机训练二分类分类器，实现来自不同社交网络的用户对关联与否的判断。

4 用户关系特征提取

4.1 基于网络表示学习的特征提取

用户关系是社交网络平台的基础特性，通过用户之间复杂的关联关系可以将独立的用户个体连接成为网络社区，也就形成了社交网络。在不同平台中，用户关系有不同的含义，主要分为关注关系和好友关系，代表相应的社交网络拓扑图是有向图或者无向图，例如，新浪微博是一种典型的有向社交网络，用户之间建立的关联是单方面的关注关系；facebook 是一种典型的无向社交网络，用户之间建立的关联是需要双方面确认的好友关系。

为了实现基于用户关系的用户身份关联，需要将用户关系转换成为下游关联模型可以读入的特征。利用网络表示学习方法将节点表示为低维稠密的向量，可以作为后续关联分析模型的输入，如图 3 所示。

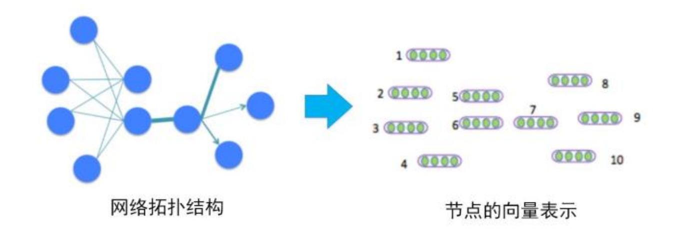


图 3 网络表示学习

Fig. 3 Network representation learning

本文利用目前流行的网络表示学习方法 LINE 算法[17]，实现用户节点的低维稠密向量的表示，既可以面向有向图，也可以面向无向图。同时，由于本文的研究对象的不是传统网络结构，而是跨社交网络这样的异构信息网络，如问题分析中的图 1 所示，研究对象网络源自两个不同的社交网络平台，所以节点类型有两种，节点之间的关系也有两种。传统的网络表示学习方法面向单一社交网络平台内部，不能反映跨社交网络的用户关系，所以为了充分利用跨社交网络先验关联关系，提高网络表示学习对异构信息网络的适用性，本文提出了 CSN_LINE 算法。

4.2　传统的 LINE 算法

传统 LINE 算法定义了一阶亲密度和二阶亲密。

一阶亲密度代表两个节点之间的直接亲密程度，对于通过边（i，j）相连的节点 v_i 和 v_j，如果边（i，j）的权重为 w_{ij}，则 w_{ij} 可以体现节点 v_i 和 v_j 的一阶亲密度，若节点 v_i 和 v_j 之间没有连接的边，则其一阶亲密度为 0。

经过节点 v_i 和 v_j 之间的实际概率和经验概率建模，为了使网络表示学习的结果向量保留一阶亲密度的特性，则需要最小化一阶亲密度实际概率分布和经验概率分布的差异，可以通过 KL 散度来衡量两个概率分布的距离，最终，一阶亲密度的目标函数如式（1）所示。

$$O_1 = -\sum_{(i,j)\in E} w_{ij}\log p_1(v_i, v_j) \quad (1)$$

其中:P_1（v_i，v_j）为一阶亲密度的实际概率建模。

二阶亲密度表示两个节点之间的间接亲密程度，通过两个节点的相同邻居节点的数量来衡量，假设如果两个节点有许多相同的邻居节点，那么这两个节点也会更加亲近。如果没有任何节点同时连接 v_i 和 v_j，那么 v_i 和 v_j 的二阶亲密度为 0。

与一阶亲密度的 LINE 算法同理，可以通过 KL 散度来衡量两个概率分布的距离，二阶亲密度的目标函数如公式（2）所示。

$$O_2 = -\sum_{(i,j)\in E} w_{ij}\log p_2(v_j|v_i) \quad (2)$$

其中:P_2（$v_j|v_i$）为二阶亲密度的实际概率建模。

通过最小化式（1）（2）的目标函数，可以生成保留了节点之间的一阶亲密度和二阶亲密度特性的向量表示。

4.3　CSN_LINE 算法

结合跨社交网络用户身份关联的应用场景，为了充分利用不同社交网络之间的关联关系，实现异构信息网络的网络表示学习，本文提出融合先验关联关系的一阶亲密度 LINE 算法。

针对跨社交网络关联的边（v，u），节点 v 和 u 分别来自不同的社交网络平台。节点间的实际概率分布如式（3）所示。

$$p_3(v,u) = \frac{1}{1+\exp(-\vec{v}^T\cdot\vec{u})} \quad (3)$$

由于边（v，u）表示先验关联关系，也就是已知的属于同一个自然人的账号对之前的关系，主观而言，该类型的边的重要程度应该远高于单一平台内部用户关系的边，故在其经验概率分布的公式中添加一项调节参数 δ，经验概率如式（4）所示。

$$\hat{p}_3(v,u) = \frac{\delta w_{vu}}{W} \quad (4)$$

其中:W 表示网络中所有边的权重之和，δ 为调节参数。对于边无权重的网络而言，W 即表示边的数。如果两个节点之间有连接，则 w_{vu}=1，如果两个节点之间没有连接，则 w_{vu}=0。

同理，融合先验关联关系的一阶亲密度的目标函数如式（5）所示。

$$O_3 = -\sum_{(v,u)\in M} \delta w_{vu}\log p_3(v,u) \quad (5)$$

通过最小化式（5）目标函数，可以使得节点的向量表示也包含跨社交网络的先验关联关系的特征。

本文提出的 CSN_LINE 算法，针对式（1）（2）（5）的目标函数进行优化，利用随机梯度下降方法，学习节点的向量表示，最终可得到两个不同社交网络的所有节点的低维向量，该向量不仅体现了平台内部的关联关系，也体现了平台之间的关联关系，可作为节点的用户关系特征，通过网络表示学习完成了用户关系特征的提取。

5　基于多层感知机的用户身份关联模型

为了确定来自不同社交网络平台的两个用户是否属于同一自然人，可以将该问题转换成为二分类问题，输入为两个不同社交网络平台的用户特征向量，输出的分类结果为 1 或-1,其中 1 代表两个用户属于同一自然人，-1 代表不属于同一自然人。

在跨社交网络用户身份关联模型中，选用多层感知机 MLP 作为分类器，由于目标为二分类，所以输出层设置为两个神经元。将待关联的第一个社交网络平台的用户节点向量和第二个社交网络平台的用户节点向量进行拼接，成为一个长向量，作为多层感知机的输入，输入的神经元个数即为拼接向量的维数。如图 4 所示，其中隐层的层数和神经元数量没有实际含义。

将一批已知的属于同一自然人的账号对拼接向量作为正样本，不属于同一自然人的账号对拼接向量作为负样本，对多层感知机网络进行训练，训练好的分类器即可作为用户身份关联模型。

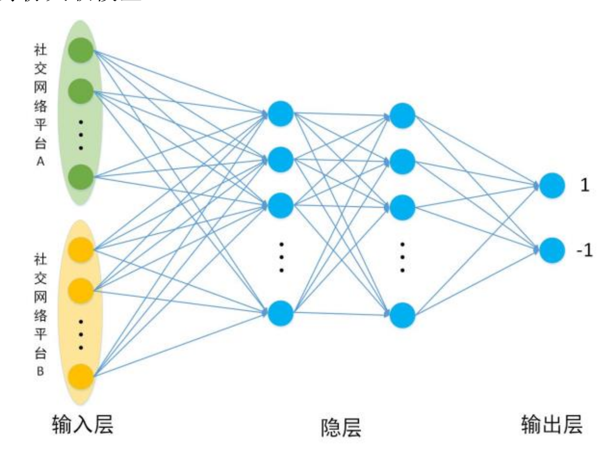


图 4　用户身份关联的多层感知机网络

Fig. 4　Multilayer perceptron network with user identity association

6　实验及结果分析

6.1　数据集描述

搜集跨社交网络平台用户身份关联的数据集是比较困难的工作，由于隐私保护的原因，几乎很难通过用户的隐私数据（手机号、邮箱号等）来确定属于同一自然人的不同账户。Veiga 等人[18]提出了一种数据收集的方法，利用用户自己发布的内容寻找跨社交网络平台线索，例如用户在某一个平台中公布了另一平台个人页面的链接，这种数据搜集方法成功运用在 Twitter、Instagram 和 Foursquare 三个境外社交网络平台上。

本文参考 Veiga 的方法，以豆瓣和新浪微博两个社交网络平台为研究对象，获取了大规模的用户关系数据，数据情况如表 1 所示。

表 1　用户关系数据情况

Table 1　User relationship data

平台\属性	节点数	连接数/边数
豆瓣	2 046 509	6 493 150
新浪微博	788 524	4 412 187

上述数据中，跨社交网络关联用户对共 14 457 对，即存

在 14 457 个豆瓣账号和 14 457 个新浪微博账号分别属于 14 457 个不同自然人。

因此，在二分类分类器的训练中，正样本为 14 457 对先验关联节点的网络表示向量的拼接，其标签为 1，同时随机选择 14 457 对不属于同一自然人的账号对，将其网络表示向量的拼接作为负样本，其标签为-1。分类器的训练样本数据共 28914 条。

6.2　评价指标

本文采用标准的评价参数进行效果评估，包括准确率（precision）、召回率（recall）、F1 值和正确率（accuracy），分别表示为 P、R、F1、Acc，计算方法如式（9）~（12）所示。

$$P=\frac{tp}{tp+fp} \tag{6}$$

$$R=\frac{tp}{tp+fn} \tag{7}$$

$$F1=\frac{2PR}{P+R} \tag{8}$$

$$Acc=\frac{tp+tn}{tp+tn+fp+fn} \tag{9}$$

其中：tp 表示正确预测为正样本的数量，fp 表示错误预测为正样本的数量，tn 表示正确预测为负样本的数量，fn 表示错误预测为负样本的数量。

6.3　常用网络表示学习算法的效果对比

为了充分利用用户关系特征进行跨社交网络用户身份关联，选择一种最适合本应用场景的网络表示学习方法十分重要，所以本实验实现了常用的网络表示学习算法 Deepwalk、LINE、Node2vec，其中包括 LINE 的三种不同模式：基于一阶亲密度、基于二阶亲密度、基于一阶和二阶亲密度。在本实验中，统一将节点的网络表示向量维度设定为 50，Node2vec 算法的随机游走参数设置为 p=0.25，q=0.25。

同时，为了衡量不同网络表示学习算法下的关联效果，经过对比实验，多层感知机的隐层设置为 2 层，每层的神经元个数为 200 个，多层感知机通过机器学习模块 scikit-learn 实现。在训练基于多层感知机的用户身份关联模型时，将 70%的数据集作为训练集，剩下 30%的数据集作为测试集。实验结果如表 2 所示。

表 2　常用网络表示学习算法效果对比 1

Table 2　Comparison 1 of effect of common network representation learning algorithm

算法	P	R	F1	Acc
Deepwalk	0.4853	0.5258	0.5047	0.4896
Node2vec	0.4834	0.5230	0.5024	0.4876
LINE（order1）	0.7191	0.7273	0.7232	0.7246
LINE（order2）	0.7596	0.9184	0.8315	0.8159
LINE（order1+2）	0.7736	0.9135	0.8378	0.8250

LINE 算法和其他两种算法是基于完全不同的网络表示策略，LINE 算法是通过优化亲密度目标函数来生成节点的向量，而 Deepwalk、Node2vec 都是通过随机游走的方式得出节点序列，然后利用类似 word2vec 的神经网络进行训练，得到节点的向量。从实验结果可以知，相对其他网络表示学习算法，当使用基于一阶和二阶亲密度的 LINE 算法时，跨社交网络用户身份关联效果最好。

6.4　CSN_LINE 算法的效果验证

本文改进了传统的 LINE 算法，基于融合先验关联关系的一阶亲密度，增加了第三个目标函数，实现了异构信息网络的网络表示学习，通过先验关联关系将两个不同的社交网络平台整合起来。

为了证明 CSN_LINE 算法在跨社交网络用户身份关联应用场景下的效果，本文设置了对比实验，在对比实验中可以进行调整的参数有两项。第一项为式（5）中的调节参数 δ，通过对比可以选择本实验中合适的调节参数值，同时证明本文提出的 CSN_LINE 算法的有效性。第二项为先验关联关系的数量，由于训练集为数据集的 70%，故训练集中有 10120 条先验关联关系，实验中可以设置不同数量的先验关联关系，以证明融合先验关联关系的一阶亲密度对用户身份关联效果的贡献。

在本实验中，调节参数 δ 值分别设置为 0、3、5、7、9，其中 0 即代表不使用融合先验关联关系的一阶亲密度。使用训练集中全部的先验关联关系，共 10120 条。对比实验结果如表 3 所示。

表 3　常用网络表示学习算法效果对比 2

Table 3　Comparison 2 of the effect of common network representation Learning algorithm

δ 值	P	R	F1	Acc
0	0.7736	0.9135	0.8178	0.8250
3	0.7975	0.9065	0.8485	0.8399
5	0.8058	0.9135	0.8563	0.8483
7	0.8064	0.9103	0.8552	0.8475
9	0.8116	0.9107	0.8583	0.8513

从实验结果分析，当调节参数 δ 值设置为 5、7、9 时，用户身份关联的效果保持在相对较高的水平；当调节参数 δ 值设置为 3 时，用户身份关联效果中等；当调节参数 δ 值设置为 0 时，用户身份关联效果相对最差。同时，相比于传统 LINE 算法（ δ 值为 0），基于 CSN_LINE 算法（ δ 值为 3、5、7、9）进行用户身份关联效果更佳，证明了本文改进的 LINE 算法的有效性。

为了证明融合先验关联关系的一阶亲密度对跨社交网络用户身份关联效果的贡献，本文进一步设置了对比实验，在调节参数 δ 值设置为 5 的情况下，分别融合训练集中 25%、50%、75%、100%的先验关联关系进行网络表示学习，实验结果如图 5 所示。

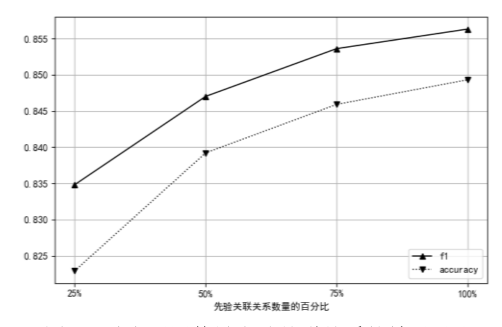


图 5　融合不同数量先验关联关系的效果对比

Fig. 5　Comparison of effects of fusing different numbers of priori correlations

从图 5 中分析可知，先验关联关系的数量越多，综合指标 F1 值和正确率越高，基于用户关系的跨社交网络用户身份关联效果越好，证明本文 CSN_LINE 算法中，先验关联关系对用户身份关联效果是有贡献的，基于融合先验关联关系的一阶亲密度进行网络表示学习能提高用户身份关联的效果。

6.5　方法对比

将本文提出的用户身份关联方法与其他两种具有代表

性的方法进行对比，证明本文方法的有效性。

第一种典型的方法是基于共同邻居节点的数量进行用户身份关联，属于无监督学习方法。核心思想是面向两个不同平台的用户，若其邻居节点中存在许多跨平台关联的节点对，则这两个用户也很可能属于同一个自然人。Zhong 等人[19]提出的 CoLink 方法中，针对两个不同平台的用户节点，计算其邻居节点中属于跨平台关联的节点对数量，然后与指定阈值进行对比，若超过阈值则判定这两个用户属于同一自然人。Sun 等人[4]提出的方法和齐林峰[20]提出的方法在用户关系方面也使用了类似的策略。

第二种典型的方法是利用网络表示学习对节点进行低维向量表示，然后利用机器学习算法进行用户身份关联模型的训练，属于有监督学习方法。Man 等人[12]提出的 PALE 方法，利用 LINE 算法的一阶亲密度进行节点表示，然后通过单隐层的多层感知机作为关联功能模块。

在本文的数据集中进行对比实验，实验结果如图 6 所示。

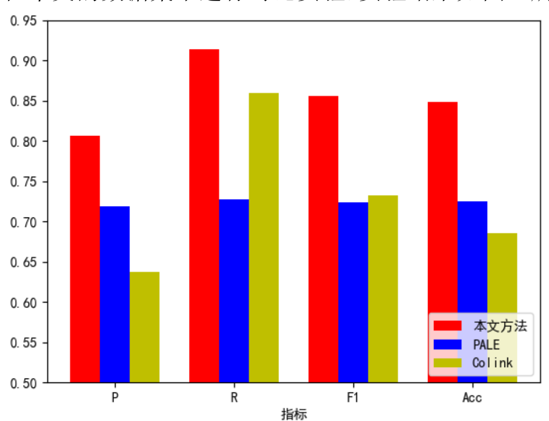


图 6 不同方法的效果对比

Fig. 6 Comparison of effects of different methods

从图 6 中可知，本文提出的基于用户关系的跨社交网络身份关联方法在 4 个评价指标下均优于 Colink 方法和 PALE 方法。具体原因在于以 Colink 为代表的基于共同邻居节点的数量进行用户身份关联的方法，特征维度过于单一，且阈值型的判断模式存在很大局限性,对相邻节点的局部差异不敏感，且 Colink 方法的准确率很低但召回率较高，其原因是真实属于同一自然人的账号对通常其邻居节点中属于跨平台关联的节点对数量较多，但不属于同一自然人的账号对的邻居节点中属于跨平台关联的节点对数量并没有明显的统计规律，所以该方法会出现较多的误判。在以 PALE 为代表的利用网络表示学习方法进行用户身份关联的方法中，由于传统的网络表示方法不擅长于处理异构信息网络，节点的向量表示不能包含跨社交网络关联关系，故关联效果不如本文提出的方法。

7 结束语

为实现社交网络平台中多源异构数据的融合，关联同一个自然人在不同平台的账号至关重要。本文提出了一种基于用户关系的跨社交网络用户身份关联方法，一方面利用网络表示学习方法将用户关系特征通过低维稠密的向量进行表示，另一方面针对本问题中跨社交网络的此类异构信息网络，通过融合先验关联关系改进了传统网络表示学习 LINE 算法，最后基于多层感知机构建了用户身份关联模型。面向大规模的实际用户关系数据，对本文提出的方法进行了训练和测试，充分证明了该方法的效果,可以应用于社交网络用户数据融合业务。

在下一步的工作中，考虑搜集更全面的用户数据，结合用户属性、用户行为、用户关系三个主要的方面，设计综合多维度特征的用户身份关联模型，进一步提高用户身份关联方法的准确性和适用性。

参考文献：

[1] 刘东，吴泉源，韩伟红，等. 基于用户名特征的用户身份同一性判定方法 [J]. 计算机学报，2015, 38(10): 2028-2040. (Liu Dong, Wu Quanyuan, Han Weihong, *et al.* User Identification across multiple websites based on username features [J]. Chinese Journal of Computers, 2015, 38(10): 2028-2040.)

[2] Zafarani R, Tang L, Liu H. User identification across social media [J]. ACM Trans on Knowledge Discovery from Data, 2015, 10(2): 1602-1630.

[3] 吴铮,于洪涛,刘树新,等.基于信息熵的跨社交网络用户身份识别方法 [J]. 计算机应用, 2017,37(8): 2374-2380. (Wu Zheng, Yu Hongtao, Liu Shuxin, *et al.* User identification across multiple social networks based on information entropy [J]. Journal of Computer Applications, 2017, 37 (8): 2374-2380.)

[4] Sun Song, Li Qiudan, Yan Peng, *et al.* Mapping users across social media platforms by integrating text and structure information [C]// Proc of IEEE International Conference on Intelligence and Security Informatics. 2017: 113-118.

[5] Vosoughi S, Zhou H, Roy D. Digital Stylometry: Linking Profiles Across Social Networks [C]//Proc of International Conference on Social Informatics. 2015: 164-177.

[6] Sha Ying, Liang Qi, Zheng Kaijiang. Matching user accounts across social networks based on users message [J]. Procedia Computer Science, 2016, 80: 2423-2427.

[7] Li Yongjun, Zhang Zhen, Peng You, *et al.* Matching user accounts based on user generated content across social networks [J]. Future Generation Computer Systems, 2018, 83: 104-115.

[8] Kong Xiangnan, Zhang Jiawei, Yu PhilipS. Inferring anchor links across multiple heterogeneous social networks [C]//Proc of ACM International Conference on Conference on Information and Knowledge Management. New York:ACM Press,2013: 179-188.

[9] Zhang Jiawei, Kong Xiangnan, Yu P S. Transferring heterogeneous links across location-based social networks [C]//Proc of ACM International Conference on Web Search and Data Mining.New York:ACM Press, 2014: 303-312.

[10] Liu Siyuan, Wang Shuhui, Zhu Feida, *et al.* HYDRA: large-scale social identity linkage via heterogeneous behavior modeling [C]//Proc of ACM SIGMOD International Conference on Management of Data. New York:ACM Press,2014: 51-62.

[11] Tan Shulong, Guan Ziyu, Cai Deng, *et al.* Mapping users across networks by Manifold Alignment on Hypergraph [C]//Proc of the 28th AAAI Conference on Artificial Intelligence. 2014: 159-165.

[12] Man Tong, Shen Huawei, Liu Shenghua, *et al.* Predict anchor links across social networks via an embedding approach [C]// Proc of International Joint Conference on Artificial Intelligence. 2016: 1823-1829.

[13] Feng Shuo, Shen Derong, Kou Yue, *et al.* Anchor link prediction using topological information in social networks [C]// Proc of the 17th International Conference on Web-Age Information Management. [S.I.]:Springer International Publishing. 2016: 338-352.

[14] Zhang Yutao, Tang Jie, Yang Zhilin, *et al.* COSNET: connecting

heterogeneous social networks with local and global consistency [C]// Proc of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.New York: ACM Press, 2016: 1485-1494.

[15] Zhou Xiaoping, Liang Xun, Zhang Haiyan, *et al.* Cross-platform identification of anonymous identical users in multiple social media networks [J]. IEEE Trans on Knowledge and Data Engineering, 2016, 28(2): 411-424.

[16] 汪潜,申德荣,冯朔,等.一种全视角特征结合众包的跨社交网络用户识别 [J].软件学报,2018,29(3): 811-823. (Wang Qian, Shen Derong, Feng Shuo, *et al.* Identifying users across social networks based on global view features with crowdsourcing [J]. Journal of Software, 2018,29(3): 811-823.)

[17] Tang Jian, Qu Meng, Wang Mingzhe, *et al.* LINE: large-scale information network embedding [C]// Proc of the 24th International Conference on World Wide Web. New York: ACM Press,2015: 1067-1077.

[18] Veiga M H, Eickhoff C. A Cross-Platform Collection of Social Network Profiles [C]//Proc of International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2016: 665-668.

[19] Zhong Zexuan, CaoYong, Guo Mu, *et al*. CoLink: An Unsupervised Framework for User Identity Linkage [C]// Proc of AAAI Conference on Artificial Intelligence. 2018.

[20] 齐林峰.利用实体解析的跨社交媒体同一用户识别 [J]. 图书情报工作,2017, 61(6): 107-114. (Qi Linfeng. The identity of the same user with cross-social media based on entity resolution [J]. Library and information service, 2017, 61(6): 107-114.)